

A NRTAC Study

2010

The Practices of Successful High-Implementing Reading First Schools

*A Study Based on Reading First Programs in
Alabama, California, Illinois, Tennessee, and Wisconsin*

NATIONAL READING
TECHNICAL ASSISTANCE CENTER



The Practices of Successful High-Implementing Reading First Schools

*A Study Based on Reading First Programs in
Alabama, California, Illinois, Tennessee, and Wisconsin*

This report was compiled by Sarah Sakyo and Corinne Eisenhart with assistance from Alice Furry and Kathleen Cooper of the National Reading Technical Assistance Center. The Practices of Successful High-Implementing Reading First Schools is one in a series of studies on topics related to the implementation of the Reading First initiative across the country.

Acknowledgments

This report was compiled by the National Reading Technical Assistance Center. Its authors are grateful to the numerous Reading First practitioners in the selected states who contributed to this report, including: Judy Stone, Alabama Reading First Initiative (ARFI), Alabama Department of Education, Montgomery, AL; Sharon Johnson, Gerri Muller, Carrie Roberts, California Department of Education, Sacramento, CA; Mark Moulton, Educational Data Systems, Morgan Hill, CA; Geri Marion, Maureen Ritchel, Illinois Department of Education, Springfield, IL; James Herman, Tennessee Department of Education, Nashville, TN; Greg Doyle, Wisconsin Department of Education, Madison, WI.

Editorial and production services provided by Elizabeth Goldman, C. Ralph Adler, and Robert Kozman of RMC Research Corporation. Original design by Lisa T. Noonis.

2010

NATIONAL READING
TECHNICAL ASSISTANCE CENTER



This document was developed by the National Reading Technical Assistance Center. The document was produced under U.S. Department of Education Contract No. ED-08-CO-0123 with RMC Research Corporation. The views expressed herein do not necessarily represent the policies of the U.S. Department of Education. No official endorsement by the U.S. Department of Education of any product, commodity, service, or enterprise is intended or should be inferred.

<http://www.ed.gov/programs/readingfirst/support/index.html>

2010

Contents

Executive summary	1
Introduction	3
What is Reading First?	3
Key questions	3
The literature on implementation	5
Defining implementation	5
Implementation frameworks	6
Doing it right: Fidelity of implementation	7
Determining high implementation for Reading First	10
Summary of key literature and state findings	15
The implementation practices of successful high-implementing Reading First schools	17
The relationship between increased student reading achievement and high implementation in Reading First schools	21
Determining success in student reading achievement for Reading First	21
The relationship between length of time in high-implementing Reading First schools and high student reading performance	25
Concluding remarks	29
Appendix	31
References	35



Executive summary

As part of the No Child Left Behind legislation, Reading First provided guidance and funding to Title 1 schools to improve students' reading outcomes, based on the low numbers of students who were meeting performance standards in reading across the country. By focusing on the earliest grades, Reading First intended to provide a comprehensive and scientifically based instructional model for teaching reading, with a goal of having every child reading on or above grade level by the end of third grade. The Reading First model was defined by specific components and required a commitment from school-based K–3 staff members to implement these components. This model has challenged educators to change the way they think about teaching and learning, and the systems that support student learning, in order to increase the probability of students' reaching the goal.

Although many people may assume that every school that implemented Reading First did so with fidelity and achieved the results intended by the program developers, this is not the case. In order to draw some generalizations about what may contribute to success, this study focuses on only those Reading First schools that have high fidelity in implementation of the Reading First model and evidence of raising student reading outcomes. Important questions emerged about successful high-implementing Reading First schools:

- How is success determined?
- How is high implementation defined and measured?
- Which implementation practices have the greatest impact?
- Is there a relationship between increased student reading achievement and high implementation?
- Is there a relationship between the length of time in Reading First and success in implementation and performance?

This study examines these complex questions. It endeavors to share documented findings of selected state evaluation reports as they pertain to successful high-implementing Reading First schools. In addition, the study reviews key findings from the literature relevant to implementation and perceptions of how successful high-implementing Reading First schools achieved their status.

This study synthesizes information from several sources, including a careful review of selected states' external evaluation reports, relevant literature and research pertaining to implementation, documents obtained from state education agencies pertaining to implementation of Reading First, and personal interviews with selected state literacy leaders and spokespersons.

Furthermore, this study attempts to summarize the current findings pertaining to successful high-implementing Reading First schools by investigating how states have operated Reading First and measured high implementation successfully.

Results indicate that leadership, teacher self-efficacy in teaching reading, professional development, and use of interventions seem to be the greatest factors influencing high implementation. In general, high implementation

seems to be a predictor of increases in student reading achievement. This study also calls for agreement in defining key terms related to implementation, further development in psychometric measures for evaluating implementation, and additional research into the relationship between length of time in Reading First and increased student reading achievement.



Introduction

Many external school reform efforts throughout the years have been developed with the goal of improving schools and increasing student achievement (e.g., the Reading Excellence Act, 1998). Many of these efforts have focused on changing instructional practices, providing educators with professional development, altering procedures for how schools function, and increasing communication and collaboration among staff. Among these efforts, Reading First was the largest federally funded early reading initiative to date.

A key distinction between Reading First and past early reading initiatives was its direct focus on ensuring that every child could read at or above grade level by the end of third grade. Structured as a competitive grant, Reading First funds were intended to serve as “seed” money to allow local educational agencies to develop an effective, scientifically based reading model that could flourish and be sustained over time. State educational agencies and local educational agencies were required to conduct external and internal evaluations and report student reading achievement.

The first step in examining successful high-implementing Reading First schools is to identify the critical aspects of the initiative to understand what Reading First is and is not. Once that understanding has been established, determinations about implementation can be made.

What is Reading First?

The Reading First Initiative was based on the science of beginning reading. It was intended to:

- Establish scientifically based K–3 reading programs
- Provide professional development so teachers have effective tools to help their students learn to read
- Select and administer screening, diagnostic, and classroom-based instructional reading assessments
- Select and implement effective instructional materials, programs, learning systems, and strategies that are scientifically based and proven to prevent or remediate reading failure (U.S. Department of Education, 2002, pp. 2–5)

The key components of a high-quality reading program include a coherent instructional design, explicit instructional strategies, coordinated instructional sequences, ample practice opportunities, aligned student materials, a 90-minute reading block, small flexible instructional groupings (with movement based on ongoing assessment), curricula for differentiating instruction, active student engagement, clearly articulated academic goals, effective classroom management, and high levels of time on task (U.S. Department of Education, 2002, p. 6).

Key questions

How is success determined? How is high implementation defined? Which implementation practices have the greatest impact? Is there a relationship between increased student reading achievement and high implementation? Is there a relationship between length of time in Reading First and success in implementation and performance? This study seeks to provide insight into these questions.

The study uses a multiple-methods approach to document evidence on successful high-implementing Reading First schools. Methods used include a synthesis of existing research, a review of relevant documents pertaining to implementation, and interviews conducted with state literacy leaders.

The study's first step entailed a review of all of the 52 available 2007 state and territory external evaluation report summaries of Reading First. The complete reports are available online on the U. S. Department of Education website at <http://www.ed.gov/programs/readingfirst/evaluationreports/index.html>. Reports were selected for this research synthesis based on the extent of information available in three areas:

- The implementation practices of successful high-implementing Reading First schools
- The relationship between student reading achievement and high implementation in Reading First schools
- The relationship between length of time in high-implementing Reading First schools and high student reading performance

Of the 52 evaluation report summaries reviewed, five were deemed to have enough information on one to three areas to warrant inclusion in the study. These five state external evaluation reports were examined in depth. California's 2008 external evaluation report was made public at the time of this study and was therefore used in lieu of the 2007 report.

As a complement to these report summaries, state documents relating to implementation of Reading First were obtained and incorporated throughout the study where appropriate. These documents were selected to illustrate the practices of selected states related to successful high-implementing Reading First schools.

Individual interviews were conducted with state literacy leaders from geographic regions throughout the nation. Information collected in the interviews offers additional insight to explain the features of these successful high-implementing Reading First schools.

The literature synthesis is intended to define key terminology and features necessary to understanding and interpreting the study questions being investigated.



The literature on implementation

Research on the implementation of external school reform initiatives is minimal and even more scant for early reading initiatives. Most of the available implementation information derives from the health and human services field. In that arena, a number of articles and studies address the implementation of various interventions (Fixsen et al., 2005). In particular, the health and human services literature addresses topics such as defining implementation, determining levels of implementation, and measuring and evaluating implementation. Many of these issues are relevant to this study and will be discussed.

Defining implementation

Defining the term “implementation” is challenging. Many authors use the terms “diffusion” or “dissemination” synonymously with implementation, while others use these terms for concepts that are distinctly different from implementation. One implementation definition is “the extent to which each component of the reform model is used” (Desimone, 2002). Similarly, some authors define implementation simply as “to use,” “carry out” or “put into effect.”

The definition of implementation as used in this study:

“A specific set of activities designed to put into practice an activity or program of known dimensions”
(Fixsen et al., 2005).

Furthermore, the related terms “initiative,” “innovation,” “intervention,” and “program” are sometimes used interchangeably and sometimes refer to as separate ideas when describing implementation. For the purposes of this study, these terms will be used interchangeably. Table 1 provides implementation definitions for the states included in this study.

Table 1: Implementation definitions

State	Definitions
CA	Fidelity of implementation is defined as “the degree to which an intervention [or program] is implemented as planned” (Haager et al., 2008, Chap. 3, p. 72).
IL	“The extent to which the Illinois Reading First (ILRF) activities were fully implemented and perceived to be useful components of an on-going effective reading program” (MGT of America, Inc., 2007, Chap. 3, p. 4). “The expectations for districts implementing Reading First are specific to indicators related to four categories (assessments, instructional program, instructional leadership, and professional development)” (MGT of America, Inc., 2007, pp. 4–5).
AL	“... implementation [is] how skillfully and faithfully schools carry out the elements of Reading First—what the local grant says they will do and how they actually do it” (State Literacy Leader, 2009).
WI	“The extent to which instruction is based on scientifically based reading research and the five essential components of reading, the significant use of student assessments and data, and how professional development influenced change in classroom reading instruction” (Miller et al., 2007, p. 2).
TN	“Following the procedures for implementation that were written into the grant” (Grenhan, Heegel, Boyraz, & Huang, 2007, p. 8).

The lack of consensus in defining “implementation” reflects the scarcity of research in this area. Nevertheless, there is agreement within the research literature that implementation is a process, not an event. It requires organizational change and an understanding that the conditions surrounding an initiative’s use affect the quality of implementation and subsequent results (Fixsen et al., 2005; Guldbrandsson, 2008). The California external evaluator adds that implementation should be defined by what happens at the school site.

Implementation frameworks

There is agreement in the literature that the process of implementation is complex and can be affected by numerous interrelated issues at the individual and organizational levels (Durlak & DuPre, 2008; Fixsen et al., 2005). Currently, the two prominent implementation frameworks come from the fields of education and health and human services. Though both frameworks use different terminology and have different theoretical bases, they share noteworthy similarities. Having a framework for thinking about implementation helps bring to life the idea that implementation is a process and needs to be understood, monitored, and evaluated as such.

The Concerns-Based Adoption Model (CBAM) is an educational framework for understanding implementation and how people respond to change. It comprises three parts: *Stages of Concern*, *Levels of Use*, and *Innovation Configurations* (Hall & Hord, 2006). *Stages of Concern* relates to how aware individuals are of a need for change and their considerations about an innovation, on a continuum from no awareness to high recognition of a need for change. The *Levels of Use* continuum represents how involved and experienced individuals are in using the innovation, from novice to expert, in moving through implementation stages, from nonuse through renewal.

Innovation Configurations seek to identify the critical components of an innovation so that observers can plainly see them in daily practice.

The National Implementation Research Network (NIRN) implementation framework for health and human service settings offers a different way to think about what is needed to make change happen. The NIRN focuses on the implementation of evidence-based practices in fields such as medical, human services, and recently, education. The NIRN implementation framework begins with exploration of an innovation, moves through installation and initial implementation, progresses to full implementation and innovation, and finally ends with sustainability. During the implementation process, staff behavior is developed and supported by *implementation drivers* that interact and integrate their influence on the culture and organizational functioning of the school.

Applied to education, implementation drivers are:

Variables present in every school system, which are controlled by staff and which can be used as tools or strategies to support the implementation or continuation of evidence-based practices (Fixsen & Paine, 2009).

There are clear common elements between the CBAM and NIRN implementation frameworks. First, implementation is understood to be non-linear; each stage interacts with one another. In other words, not all components of an initiative are simultaneously implemented. For example, when faced with a potential barrier such as a loss of a principal, a school may recede from full implementation of Reading First to initial implementation until the barrier is overcome.

Second, similarities at high levels of implementation indicate that the staff members are skilled, competent, and able to modify practices in purposeful ways. High levels of implementation are characterized by an ability to adapt practices to particular contexts or needs while maintaining true to the initiative's critical components.

Another common feature is that new educational practices are integrated into existing school and district systems; they no longer exist in isolation, but become part of the school culture. Finally, high implementers go through ongoing phases of renewal—success creates the energy, motivation, and momentum needed to continue.

Doing it right: Fidelity of implementation

Implementation processes need to be purposeful and described in sufficient detail so that independent observers can detect the presence and strength of the “specific set of activities” related to implementation (Fixsen et al., 2005). The most common description of fidelity of implementation is “the degree to which an intervention or program is delivered as intended by the program developers” (Carroll et al., 2007). Generally speaking, definitions of fidelity of implementation in the health and human service field tend to focus on implementation of program elements (O'Donnell, 2008). Applied to K–12 core curriculum interventions, definitions of fidelity of implementation often reference instructional quality (O'Donnell, 2008). For example, Loucks (1983) describes fidelity of implementation as “related to the amount of change that occur(s) in the teacher's practice.” The *Reading First Implementation Evaluation Interim Report* defines fidelity of implementation as “implementing the major elements of the program as intended by the legislation” (U.S. Department of Education 2006, p. 3).

Some emerging implementation literature and research distinguishes between implementation and program elements, yet it is clear that both should be taken into account when defining *fidelity of implementation* (Carroll et al., 2007; Durlak & DuPre, 2008; Fixsen et al., 2005; Guldbransson, 2008; Kurki, Aladejem, & Carter, 2005; O'Donnell, 2008). Both elements interact and influence each other in powerful ways. Because implementation fidelity acts as a potential intermediary between programs and their intended outcomes, it is crucial to explore what the influencing implementation elements are for Reading First (Carroll et al., 2007). A conceptual framework for *fidelity of implementation* is offered below to describe the Reading First implementation elements.

The developers of the framework (Carroll et al., 2007) describe the five most common elements of implementation fidelity, listed and defined in Table 2. Two additional elements (intervention complexity and facilitation strategies) are new and important elements that complete the picture of implementation fidelity. These additional elements appear in Table 2 as items six and seven.

The element of “adherence” is the baseline for determining implementation fidelity, while elements two through five act as potential moderators between the program and high-quality outcomes. These four elements can be split into two groups: fidelity to structure (adherence, exposure) and fidelity to process (quality of delivery, program differentiation), with the element of participant responsiveness taking on characteristics of both structure and process. The element of intervention complexity is important and should not be ignored. The authors have concluded that the specificity of a program (i.e., intervention complexity) enhances adherence (Carroll et al., 2007, Desimone, 2002).

Table 2: Seven elements of implementation fidelity

Elements	Definition	Application to Reading First
1. Adherence	A program being delivered as it was designed	The Reading First elements as described by legislation
2. Exposure	The amount of an intervention received by participants	The frequency and duration of reading instruction is as prescribed by program developers and matched to student need
3. Quality of delivery	The manner in which a staff member delivers a program	Instructional delivery using methods prescribed by the core learning system and/or applying proven methods to deliver content
4. Participant responsiveness	How far participants respond to, or are stimulated by, the intervention	Student motivation and engagement in reading-related tasks
5. Program differentiation	Identifying which elements of a program are essential, without which the program will not have its intended effect	Reading First’s theoretical base and practices that can be distinguished from other schoolwide reading reform models
6. Intervention complexity	How detailed and specific a description of an intervention is	The specific Reading First elements and program descriptions as provided in the legislation and Reading First Guidance document
7. Facilitation strategies	Approaches that optimize the level of fidelity achieved	Curricular manuals, internal and external training, monitoring and feedback from coach and principal, collaboration among staff, capacity building

The elements are interrelated in complicated ways. It is in the element of “quality of delivery” that the influence of teacher ability appears. For Reading First, the “quality of delivery” will depend on a teacher’s ability to deliver the content appropriately (i.e., explicit, systematic instruction using a gradual release of responsibility model, incorporating corrective feedback and differentiated instruction). This requires understanding student needs as determined by assessment data and classroom observations, being able to plan effective instruction based on student need, and lastly, delivering that plan in a manner that directly addresses student need. Teachers need to adjust instruction over time and monitor outcomes based on student responsiveness to the instruction. Even though the Reading First legislation requires the use of a comprehensive learning system for reading that is based on scientific reading research and reflects the five essential reading components, use of such a program does not guarantee sound instruction, nor does it necessarily lead to good reading outcomes. The comprehensive learning system is a tool to teach the reading curriculum: it does not deliver itself, but depends on a knowledgeable and effective teacher to make sound judgments about the instructional delivery based on student need. Even so, a knowledgeable and effective teacher will need resources, support, and feedback (i.e., “facilitation strategies”) to increase efficiency in attaining implementation fidelity.

This leads to considering the inclusion of *adaptation* in defining fidelity of implementation. Adaptation refers to changes made to the original program design (Durlak & DuPre, 2008). Other authors have drawn attention to adaptation as an element of fidelity of implementation (O’Donnell, 2008). Though adaptations can enhance effectiveness, high implementation must be reached before adaptations can be made successfully. This is because until high implementation is attained, implementers of an intervention such as Reading First will not realize the outcome potential. Only after achieving high implementation can staff begin to make modifications to reach the few students who are still struggling to read. Put simply, teachers not only need to know what to teach in reading, but how to translate the research into procedures that are guided and adjusted by their sound judgment and student need. Once this distinction is realized, it becomes apparent that a teacher’s content knowledge (“knowing what”), procedural knowledge (“knowing how”), and conditional knowledge (“knowing when and where”) of reading instruction affects his or her practice. When thinking about “high” implementation, teachers’ ability to make modifications in light of the goal is an important variable to take into account. That being said, adaptations should not compromise the program’s integrity or effectiveness or stray too far from the characteristics that define Reading First. Adaptations should enhance outcomes rather than undermine success.

Time and time again in the literature, fidelity of implementation is defined only by the adherence definition. The problem with such a narrow definition is that how much the program is used has little to do with the quality or weight of the evidence regarding that program (Fixsen et al., 2005). The fidelity of implementation definition presented here stresses that the circumstances that affect the implementation of distinct program elements may be just as important as the elements themselves. For this reason, each element above not only affects the innovation processes, but also the implementation processes. Implementation fidelity can be thought of as the “how” for installing the “what.” The features of implementation must be identified just as explicitly as the program components have been identified (Fixsen et al., 2005). This way, the implementation strategies that improve program fidelity will be more apparent and potentially easier to disseminate and replicate in other schools.

Other factors, such as the implementation process elements described above, also affect successful program implementation and must be part of the picture so that no single factor solely determines implementation fidelity.

"There were just so many things beyond the scope of the mere implementation of Reading First elements that made it very different [for low-implementing RF schools]."

"When you are talking about implementation, you are talking about an accumulative effect. It becomes part of the life of the school. So, when you are talking about implementation at the beginning and implementation at the end, you are talking about something that looks very different. When we talk about implementation, we need to be sure that we are not just talking about a set of rules that we followed, that we put into schools to improve reading. There are changes due to things that we don't measure or even discuss. We hold individual grade-level meetings now. Those kinds of things we learned made a difference in implementation. It is a culture change."

—State literacy leaders

A major challenge in studying fidelity of implementation for K–12 core curriculum innovations concerns the inconsistencies in how the term is defined across studies, which make it nearly impossible to measure implementation accurately (O'Donnell, 2008; Durlak & DuPre, 2008). The complexity of measuring "quality of delivery" and the other fidelity elements will be discussed later. "Doing it right" only means that a school is in compliance with the law. For this reason, the existence of the program elements alone, even if followed to the letter of the law, will not guarantee good outcomes. Before investigating how the states measure implementation, it is critical to determine what high implementation means for Reading First in these selected states.

Determining high implementation for Reading First

In a simple sense, if adherence to the program elements is achieved, then it can be said that implementation is high. Once fidelity of implementation is understood from the broader perspective noted above, however, it becomes difficult to quantify high implementation concretely. How high implementation is determined depends on the perspective. Some people view implementation from an adherence mentality, while others see it through a change framework (such as CBAM or NIRN), which incorporates implementation process features. Because implementation process features are difficult to quantify, many state Reading First external evaluations have used the adherence perspective to determine implementation levels. The advantage of this perspective is that it allows for an accurate and reliable—if one dimensional—portrait of high implementation. Therefore, some state Reading First external evaluations measure to what degree the critical Reading First components are functioning as intended by law. The closer the match of the Reading First components—as described by the program developers—to the application of those components in the school, the higher the level of implementation. It is also important that states establish criteria to accurately and consistently compare program implementation by various schools with the criteria in order to determine a level of implementation (e.g., low, moderate, high).

Table 3 uses the NIRN description of high or full implementation for an evidence-based schoolwide reading model. The description attempts to showcase how individuals might use implementation drivers at the full implementation stage to affect program implementation. It should be pointed out that some of the indicators listed below reflect practices and conditions that are beyond the scope of the critical Reading First components as determined by law. These system practices and conditions are hard to measure, yet have some of the greatest impact on successful implementation as determined by the literature.

Table 3: Full implementation model

(Fixsen & Paine, 2009)

Descriptors of full implementation
<ul style="list-style-type: none">• A majority of staff are using the new procedures according to standards to which they have been trained and coached• Program elements and supports are in place and are working at the individual and organizational levels• Changes in instruction, assessment and program support are now well under way• Outcomes cannot meaningfully be evaluated until this point• The goal is to embed the new practices and the implementation drivers into teacher practices and support systems to maintain implementation and positive results• Processes that led to this point must be continued (ongoing training, coaching, leadership, data monitoring, etc.)• The work of the implementation team (i.e., the group chosen to help put effective practices in place) is critical to getting to this stage; keeping it there is critical as the staff adjusts to changes in context (e.g., staff turnover, variation in implementation strength over time) and bridges to sustainability

If implementation is viewed as “a specific set of activities designed to put into practice an activity of known dimensions,” then two aspects are critical in determining “high” implementation: a program and the conditions surrounding its use (Fixsen et al., 2005). Accordingly, the descriptors mentioned above can be viewed as conditions that can influence Reading First program implementation and, consequently, should be operationalized so that implementers can strengthen them. Implementation fidelity may affect the relationship between the two variables of innovations and outcomes, which is the primary reason that it needs to be measured. *What implementation activities put Reading First into practice successfully? Are some implementation activities more powerful than others?* This study will attempt to provide insight into this question later in this document.

Of course, it is crucial to analyze whether a state's established criteria for determining high implementation include success in raising student reading achievement. How one state determines success may differ from another, depending on how each defines proficiency, on the expectations—if any—for percent proficient, and on whether performance on the criteria can predict student outcomes. If the critical program components are present in a school, but reading achievement results are low, the school should not be considered a "high implementer." Table 4 displays the instruments used to measure "high" implementation and Table 5 presents examples of how states quantify "high implementers." What one state designates as high implementation may not be considered high by another.

Table 4: Instruments to measure high implementation

State	Measurement instruments
CA	The Reading First Implementation Index (RFII), based on the CA "assurances," is a survey administered to teachers, coaches, and principals, varying across grades and core program type. The survey data are analyzed using a 3-Facet Rasch Model to produce an implementation measure for each school (RFII). The methodology ensures that the RFII is comparable across schools and years, regardless of variations in the mix of items (Haager et al., 2008, p. 82).
AL	Site visits using standard observation protocol Phone interviews Yearly state monitoring visit (Moscovitch, 2008).
IL	The Implementation Success Variables are based on the principal/coach interview rating, teacher interview rating, student on-task behavior observation composite, principal survey composite, coach survey composite, and teacher survey composite. The student on-task behavior observation composite was based on the student engagement rating from the Instructional Content Emphasis-Revised (ICE-R), which ranged from high to low engagement (a rating of 1 to 3). The Principal, Coach, and Teacher Survey composite variables were the sum of the survey ratings across each of four sections of the survey for each stakeholder group. The total survey composite was the sum of the three stakeholder survey composites (MGT of America, Inc., 2007, Chap. 3, p. 2-4).
WI	Classroom observations Teacher, principal, and literacy staff interviews Teacher survey (Miller et al., 2007, p. 2).
TN	Literacy Observation Tool (LOT) Intervention Observation Tool (IOT) Teacher, principal, and literacy leader questionnaires Implementation Benchmark Tool (school-developed program implementation benchmarks) (Grenhan et al., 2007, pp. 11–19).

Table 5: Definition for quantifying high implementers

State	Examples of quantifiers
CA	A school's relative score between 1 and 100 on the annual Reading First Implementation Index survey [requiring all teachers, coaches, and principals to participate] provides a comparative benchmark for designating high-implementation schools: a cut-score of 41.4 indicates the bottom boundary for the higher implementing schools and a cut-score of 36.0 indicates the upper boundary of the lower implementing schools [separated by one standard deviation]. The range between these two cut-scores, or mid-range, is not reported in the evaluation studies (Haager et al., 2008, p. 82).
AL	"We rank our schools, but not very scientifically. Schools where teachers skillfully and faithfully implement their Reading First plan are ranked as high. If some teachers do and others don't, which is the case in some schools, [this] would be a moderate implementation. If many of the teachers do not, that would be considered low implementation" (State Literacy Leader, 2009).
IL	Based on the total survey composite variable, the schools scoring in the top 25 percent of the distribution were coded as high implementing schools (MGT of America, Inc., 2007, Chap. 4, p. 42).
TN	A self-assessed rating of Phase III (Full Implementation) on the Reading First Implementation Benchmark Tool. A generic Implementation Benchmark Tool was given to each school to customize. The Implementation Benchmark Tool consists of 17 items, each containing three phase levels, related to the areas of curriculum, instruction, and organization. The five components of effective reading instruction were included in the generic model. Each statement was then elaborated by specific indicators and evidence for the implementation phases (Grenhan et al., 2007, pp. 17, 102-103, Appendix A).

In California's annual spring implementation survey on the essential elements of its Reading First program, 18,623 or 95%, of the possible respondents responded. Three out of 18 dimensions of the program are used to calculate the 19th dimension or the Reading First Implementation Index (RFII). These three dimensions are weighted as representing the critical essential elements of the program, with 70% of the index score based on School Implementation Overall (SIO) items (210/238); 20% on Overall Reading First Understanding (ORFU) items (17/238); and 10% on Teacher/Coach Professional Development (TCPD) items (11/238). A separate dimension, outside the RFII total score, is the School Implementation and Instruction (SII) score with 28 items, which has been found to be the most powerful predictor. The RFII score of SIO and ORFU are the second and third most powerful predictors of school success (Haager et al., 2008, p. 81).

Table 6 indicates the percentage of “high implementers” in states when percentage of proficiency or gain results are reported.

Table 6: Percentage of designated high implementing schools

State	Percentage of schools
CA	Approximately 18% of the schools across all program years were designated as high-implementation schools. When reported in years, 11% of the schools in year six, 20% of the schools in year five, and 17% of the schools in year four were designated as high implementation schools (Haager et al., 2008, p. 28, 31).
AL	“An estimate of 25–30% made proficiency goal based on number of students’ proficient” (State Literacy Leader, 2009).
IL	25% (top 25% of distribution on the survey composite was used to define high implementation) (State Literacy Leader, 2009).

Currently, psychometrically sound ways to define and, consequently, measure implementation are limited (Zhang, Shkolnik, & Fashola, 2005). There are two sets of issues to consider in measuring implementation, based on the extent to which the process of implementation is teased apart from the method, or innovation. *Innovation* processes and *innovation* outcomes must be clarified and distinguished from *implementation* processes and *implementation* outcomes (Guldbransson, 2008). To achieve a good end result, *both* the method and the implementation must work well.

Furthermore, the element of “quality of delivery” is more ambiguous than it seems and, therefore, more difficult to measure reliably. “Quality of delivery” can serve as a moderator between program implementation and outcomes, but this role has not been studied in the implementation literature to date. It is not hard to imagine that a teacher can be delivering instruction, but ineffectively. This reinforces the adherence-only issue, and can ultimately result in lower implementation.

“Our emphasis is evaluation of the schools, not the respondents. We tried to set up the survey as a way for respondents to describe the practices of their schools.”

—An external evaluator

The considerations above bring more depth and breadth to how implementation has traditionally been measured and would require a more complex evaluation methodology because they suggest a more comprehensive way to view implementation.

There is common agreement in the literature that it can take upwards of four years to attain full implementation (Fixsen et al. 2005; Durlak & DuPre, 2008). Therefore, evaluating program implementation should occur only once full implementation has been reached (Desimone, 2002). Evaluations that occur too early in the implementation process may produce poor results that may not be due to an ineffective program, but rather to poor or incomplete program implementation (Fixsen et al., 2005; Carroll et al., 2007). States and evaluators should be able to determine full implementation if criteria have been established to define what full implementation entails.

"A long time is needed for implementation to really hold and sustain ... some schools, the really difficult schools, need more than three years."

—A state literacy leader

What is known about the implementation of K–12 core curriculum innovations? We see two disappointing findings: (a) fidelity of implementation is rarely reported in large-scale education studies that examine effectiveness, and (b) measures of fidelity to K–12 curriculum innovations are seldom used to adjust for or interpret outcome measures (O'Donnell, 2008). In measuring the relationship between implementation and outcomes for K–12 curriculum innovations, distinguishing between fidelity to the structural components and fidelity to the processes that guide an innovation's design need to be measured separately and then each related to outcomes. This is because the strongest relationships to outcomes, as concluded from quantitative studies, appear to be caused by variations in fidelity to implementation process elements supported by the curriculum innovation (O'Donnell, 2008). In order to get an accurate portrayal of the potential relationship it is necessary to identify the Reading First implementation process elements and determine how to measure their effect on strong reading outcomes.

Unfortunately, the evaluation of implementation fidelity and the potential moderating effects of the facilitation strategies and implementation drivers on a program have not been measured in any studies to date. As a result of the differing fidelity measures within efficacy and effectiveness studies, there are still too few studies to direct researchers on how fidelity of implementation to core curriculum innovations can be measured and subsequently related to outcomes (O'Donnell, 2008; Zhang et al., 2005). The lack of evidence to support these assumptions should not, however, lessen the consideration of the potential effects. "Doing it well" connects high implementation to outcomes. It depends on high program implementation and success in raising student reading outcomes.

Summary of key literature and state findings

In review, the implementation research brings to light that:

- Consensus in defining key terminology related to the study of implementation is needed.
- The development of psychometric measures for evaluating implementation is needed.
- Implementation elements should be separated from program elements.

State findings reveal:

- The selected states define implementation by an adherence definition.
- The selected states use a variety of methodologies to measure implementation, including surveys, observations, and interviews.
- Some of the selected states applied a scientific methodology for determining high implementation while others used less scientific methods.

The implementation practices of successful high-implementing Reading First schools

What does the research say about the factors that affect implementation? Answering this question is difficult due to the large number of implementation variables deemed important. A review of quantitative and qualitative implementation data indicates a convergence of evidence on factors that affect the implementation process (Durlak & DuPre, 2008; Elias, Zins, Graczyk, & Weissberg, 2003; Fixsen et al., 2005). Eleven common factors were deemed critical to influencing implementation in four reviews (Durlak & DuPre, 2008; Fixsen et al., 2005; Greenhalgh et al., 2005; Stith et al., 2006). Table 7 outlines the eleven factors.

Table 7: Eleven factors influencing implementation

Implementation factor	Conceptual category	Reading First system variables
Positive work climate	Organizational capacity	Reading culture
Shared decision-making	Organizational capacity	Leadership
Coordination with partners	Organizational capacity	Communication
Formulation of tasks (planning, grade-level team meetings, procedures, roles and responsibilities)	Organizational capacity	Planning and evaluation
Strong leadership	Organizational capacity	Leadership
Program champions	Organizational capacity	Communication
Administrative support	Organizational capacity	Standards, curriculum, and instruction; assessment and data utilization
Staff skill proficiency	Staff characteristics	Professional staffing
Training	Support system	Professional development
Technical assistance	Support system	Professional development
Funding	District and school level factors	Resources and funding

It is clear from this table that factors related to organizational capacity and the support system are important to successful implementation. Reading First has, or relies on, systems at multiple levels, involving several different functions. Ideally, the goals and structure of each sub-system would align with the organization's primary purpose—for Reading First, improving reading outcomes. A "system" therefore, is a whole which is made up of multiple parts, each of which must carry out its designated function, and all of which must work in unison to create the intended outcome. For that reason, the system-level variables for Reading First are the *implementation drivers*.

The organizational capacity and the Reading First system variables (implementation drivers) must work together to improve program implementation and achievement. Schools must recognize the system variables and prepare strategies to navigate successfully through potential implementation barriers.

"Our knowledge base of reading went 'sky-high.'...We built an infrastructure. It is everywhere."

—A state literacy leader

The authors of a report investigating how to measure and explain the fidelity of Comprehensive School Reform (CSR) implementation concluded that principal leadership, professional learning communities, and program coordinator's assistance were more likely to make a significant difference to success in reaching higher levels of implementation (Kurki et al, 2005). Will this be true for Reading First?

What implementation factors had the greatest impact on successful high-implementing Reading First schools? Do those factors add value to what the research tells us about implementation factors influencing other innovations?

The Reading First Implementation Evaluation Final Report states that, "Only the [Reading First aligned] composite measure of teachers' use of activities for struggling readers was statistically significantly related to the probability that a school scored in the top quartile (relative to other RF and non-RF Title I schools) on its state's third-grade reading assessment" (U.S. Department of Education, 2008, Chap. 8, p. 126).

In the Alabama Outlier Study, six successful Alabama Reading First Initiative schools reveal that when the principals and reading coaches engaged in extensive discussion with teachers about individual students, the teachers became more focused on instruction to address individual student needs. In particular, data meetings at these schools emphasized student needs. Principals also became more committed to help every student read by arranging common grade-level planning time, revising the school schedule based on need, and ensuring consistency of interventions. These actions significantly distinguished successful and unsuccessful schools by the amount and quality of intervention services given to their struggling reader populations (Moscovitch, 2008, pp. 7–9).

The Alabama Best Practices Center profiled four top-performing Reading First Initiative schools based on the state outcome measure from several school years by conducting on-site observations and interviews with key staff. In all four schools, staff stated that their success was primarily due to the extensive amount of high-quality professional development received, the strong instructional leadership of the principal and coach, opportunities to communicate and collaborate about student needs, and a cultural change in the way teachers approach the teaching of reading and expectations for all students (Alabama Best Practices Center, 2004–2007).

In California, the state's Reading First Implementation Index score (RFII), its stakeholders' annual survey (for teachers, coaches, and principals), and a single dimension on Teacher Reading First Evaluations, all address how teachers judged the Reading First program at their school. This teacher score (based on four items) served as one of the most powerful predictors of school achievement.

"When teachers are positive, schools grow rapidly. And when they are negative toward the program, schools suffer [thus] teacher buy-in has a causal relationship to achievement."

—An external evaluator

Statistical tests also showed that when teachers and coaches judge the principal as a full player in the school, the effect size for predicting student achievement becomes "sufficiently larger and counter-balances the serious demographic challenges posed by high SED, EL, black, and migrant education student populations" (Haager et al., 2008, pp. 81–82).

Another California finding indicates that a separate dimension, the School Implementation and Instruction (SII) score (based on 28 items) was also a powerful predictor of successful, high-implementing schools. This SII dimension is based on items that “ask about principal support of the teachers, planning time, the pacing schedule, grade-level meetings, and the principal’s involvement in these meetings...and how well the school and principal support the teaching staff” (Haager et al., 2008, p. 81).

In Illinois, evaluators observed 49 classrooms and reported the findings from the top three schools based on the state’s standardized reading assessments (ISAT, IAA, or IMAGE). Based on case studies, they found that the activities, perceptions, and outcomes that characterized the three most successful schools were attributable to professional development; use of interventions; differentiated, data-driven, small group instruction; teacher buy-in; and an uninterrupted 90-minute literacy block. The evaluation report noted that these schools spent more time in collaborative, professional discussion centered on reading instruction; conducted more regular grade-level team meetings and staff meetings; and engaged more in formal and informal classroom observations than did lower-performing schools. These activities, in effect, influenced teachers’ positive attitudes about teaching reading and improved their diagnostic insight to student needs. School and district leadership was also recognized as a factor in schools’ success (MGT of America, Inc., 2007, Chap. 4, pp. 43–47).

“Reading First schools with high levels of problematic reading instruction that had a really strong principal who was visiting those classrooms regularly and expressed consistent support for the program did just overwhelmingly better than the same kind of school with poor leadership. I can tell you that in Reading First schools that implemented this program well, the difference in the climate in the building was profound. Strong leadership is not just for Reading First but I do not think that many Reading First schools could have done nearly as well had they not had a very strong principal.”

— A state literacy leader

Similarly, the factors that have influenced successful Reading First implementation in Wisconsin included strong leadership support at the district and school levels, professional development, protocols for monitoring classroom instruction, and open communication in support of the literacy program (Miller et al., 2007, pp. 43–45).

The information collected through the Tennessee observation tool reflects strong attainment of Reading First goals. The responses from Rounds I and II teachers, principals, and literacy leaders in Year 3 suggested that they are more comfortable with the Reading First program and can focus more of their time and attention on implementation. More than 90% of the teachers and principals in 2006–07 strongly agreed or agreed that they had a thorough understanding of Reading First. More than 95% of teachers reported that assessments, programs, and supplemental materials were based on scientifically based reading research (SBRR) and systematic and explicit instruction in phonemic awareness, phonics, vocabulary, fluency, and comprehension occurred daily in their school’s K–3 classrooms. Evidence from the Intervention Observation Tool (IOT) observations indicated that in Year 4 of the program Tier II and III interventions were firmly established and effective (Miller et al, 2007, p. 5).

Literacy leaders also reported frequent or extensive involvement in coordination and implementation of the 3-Tier Reading model and working with teachers to ensure the K–3 curriculum and programs centered on the five essential elements of reading. Teachers continued to use small- group instructional activities and learning centers more often while incorporating more fluency and phonics activities into their instruction (Grenhan et al., 2007, p. 6).

A review of the responses on the questionnaire for Round I teachers indicated that in Year 4, teachers continued to feel they had received adequate levels of professional development and support. In all items related to this topic, there continued to be an increasing frequency and strength of agreement among teachers over the four years of implementation that they had received the training and support that they needed. For example, 94%

of teachers in Year 4 strongly agreed or agreed that they had received adequate initial and ongoing professional development. Approximately 89% of teachers in Year 4, compared with 77.3% of teachers in Year 1, strongly agreed or agreed that the Literacy Leaders had provided helpful guidance and support (Miller et al., 2007, p. 4).

Table 8: Most influential implementation factors for high implementers

State	Influential factors			
	Leadership	Teacher self-efficacy	Professional development	Intervention
CA	✓	✓		
AL	✓	✓	✓	✓
IL	✓	✓	✓	
WI	✓	✓		
TN	✓	✓	✓	✓

Findings from the selected states on the implementation practices of successful high-implementing Reading First schools suggest that:

- The strength of leadership (district/school and principal/coach) in establishing positive culture and shared decision-making seems to influence implementation.
- Teachers' self-efficacy (skills and knowledge) in teaching reading, high expectations for students, and buy-in to the program seem to influence implementation.
- The quality of professional development and technical assistance in building a collaborative learning climate seems to influence implementation.
- The amount and quality of activities and focused interventions seem to influence implementation.

The relationship between increased student reading achievement and high implementation in Reading First schools

[In] the effort to improve outcomes, weak results—or fleeting results—often have more to do with the manner in which new procedures are implemented than with any other variable. (Fixsen & Paine, 2009)

Determining success in student reading achievement for Reading First

It is fair to say that most, if not all, states define success in student reading achievement for Reading First schools by meeting proficiency standards on existing state reading achievement measures. This is coupled with proficiency and/or benchmark results from required Reading First screening and outcome measures. This is the case for the five states included in this study.

In their Reading First external evaluation reports, Alabama and California include data on the relationship between increased student reading achievement and high implementation. These states were selected for inclusion in this portion of the study because they have longitudinal analyses that look at high implementation as a predictor of student reading achievement. Specific information regarding measures used to determine success for each state is outlined in Table 9.

Table 9: Student achievement measures used by selected states

State	Achievement measures
CA	Proficient or benchmark: <ul style="list-style-type: none">• California Standards Test: English-Language Arts, Grades 2 and 3• CAT-6 (Grade 3 only)• End-of-Year Oral Reading Fluency measure grades 1–3• End-of-Year Kindergarten Assessment
AL	Proficient or benchmark: <ul style="list-style-type: none">• DIBELS = 80%• Stanford 10 Achievement Test (SAT-10) = 60%• Alabama Reading and Math Test (ARMT)

In order to reveal the connection between high implementation and increased student reading achievement, selected student reading achievement data from California and Alabama are presented. Please note that the California data are aggregated and are meant to paint a global picture of the relationship between high implementation and increased student reading achievement. The Alabama data showcase school-level outcomes from the profiled schools in the journal of the Alabama Best Practices Center, 2004–2007, in order to highlight individual school success. The different profiled schools (high-implementing and high-achieving schools) are featured each year from 2003 to 2007 with identified measures to document their progress.

California

"The achievement results indicate that the more "faithfully" the school implemented Reading First, the greater the student progress."

—A state literacy leader

In California, high implementation of Reading First is a statistically significant predictor of student reading achievement on all achievement metrics; the greater the fidelity of the program, the greater the effect on achievement. "High-implementing" Reading First schools in California showed higher achievement gains than "low-implementing" Reading First schools for grades 2-4. Program effectiveness as measured by the Reading First Implementation Index survey (RFII) offers 18 dimensions on "school implementation/instruction" (i.e., school support, leadership, school culture) and "overall school implementation" that strongly predicts student reading achievement. "Overall implementation" is calculated from a large number of items with an emphasis on school-level support for teachers and teacher practices in the classrooms, as well as coaching support. Table 10 illustrates how program "effectiveness" is a function of the relationship between implementation and achievement gains.

Table 10: Grade 2 Reading Achievement in California's high- and low-implementing schools

(Haager et al., 2008, p. 31)

	California standards test for English language arts Grade 2				
	All Reading First schools	High implementation school (Avg. RFII >41.4)	Low implementation schools (Avg. RFII <36.0)	Statistical control group (RFII =25.0)	All non- Reading First elementary schools
Years in program: 6					
Number of schools	253	28	96	N/A	4,057
% Proficient and above					
2002	15.5%	14.5%	15.2%	15.5	37.7
2008	35.8%	36.6%	34.8%	33.0	51.2
Change since starting year	20.3 ^{abc}	22.1 ^{abc}	19.6 ^b	17.5	13.5

^a Significantly different ($p < 0.05$) relative to the "Statistical Control Group."

^b Significantly different ($p < 0.05$) relative to the "All Non-Reading First Elementary Schools"

^c Significantly different ($p < 0.05$) relative to the starting year, i.e., significantly different from a gain of zero.

"The more you follow the program with fidelity, the greater the student success. Studies were conducted on the achievement of English Language Learners [and] findings indicated that those in high-implementing schools made more progress than those in low-implementing schools."

—A state literacy leader

California also sought to determine the extent to which the Reading First program had specifically affected English Learners (ELs). Results showed that reading achievement gains for ELs in "high-implementing" Reading First schools were almost uniformly higher than for ELs in "low-implementing" Reading First schools and non-Reading First schools. Table 11 displays higher percentages of ELs in "high-implementing" schools outperforming ELs in "low-implementing" schools at proficient or advanced levels on the California Standards Test for English Language Arts.

Table 11: Grade 2 reading achievement in California's high- and low-implementing schools of English learners

(Haager et al., 2008, p. 135)

	Reading First schools				
		English learner students			
Years in program: 6	All Reading First schools All students	All Reading First schools	High implementation schools (Avg. RFII >41.4)	Low implementation schools (Avg. RFII <36.0)	All non-Reading First elementary schools
Number of schools	253	237	26	94	2135
% Proficient and above					
2002	15.5	11.3	9.3	12.1	17.8
2008	35.8	28.5	34.6	26.3	34.7
Change since starting year	20.4	17.2 ^b	25.3 ^{ab}	14.2 ^{ab}	16.8
Mean scale score per student					
2002	300.1	293.0	290.0	294.1	304.5
2008	330.0	321.0	330.1	317.6	329.6
Change since starting year	30.0	28.0 ^{ab}	40.1 ^{ab}	23.5 ^b	25.0

^a Significantly different ($p < 0.05$) relative to English learners in "All Non-Reading First Elementary Schools".

^b Significantly different ($p < 0.05$) relative to the starting year, i.e., significantly different from a gain of zero.

Second-grade ELs in "high-implementing" schools obtained higher achievement scores from the base year of 2002 to Cohort 1's 6th year than did ELs in "low-implementing" schools during the same period; these gains were statistically greater than those of ELs in non-Reading First schools. ELs in "high-implementing" Reading First schools often made greater gains than the student population in Reading First as a whole. When the program is faithfully implemented, schools show remarkable gains in their ability to serve their populations (Haager et al., 2008).

Alabama

Student achievement results from the top-performing Alabama Reading First Initiative schools, profiled in an Alabama journal, showed that the 2003–2004 top performing school reduced the percentage of "at risk" kindergarteners in phoneme segmentation fluency as measured by DIBELS from 40% in mid-2003 to 16% in mid-2004 (The Alabama Best Practices Center, 2004, p. 10).

The school profiled in 2004–2005 had 97% of its K-3 students at benchmark by the end of year on DIBELS and ranked 3rd among all Alabama schools in grades K–3. In addition, Stanford Achievement Test (SAT) results for grade one had 70% of students at proficiency (The Alabama Best Practices Center, 2005, pp. 10, 12, & 14).

In 2005–2006, the percentage of students at benchmark in kindergarten and first grade on DIBELS ranged from the high 70's to 100%. On the Alabama Reading and Math Test (ARMT), 97% of the third graders met or exceeded the academic content standards compared with 84% of all Alabama students (The Alabama Best Practices Center, 2006, pp. 5–7).

The 2006–2007 profiled school outperformed its comparable better-funded, lower SES or higher poverty schools. More than 96% of third graders scored at or above proficiency in reading on the Alabama Reading and

Math Test (ARMT). Table 12 shows the profiled school's reading achievement results for 2007 compared with all ARFI and all Alabama schools (The Alabama Best Practices Center, 2007, pp. 6–7).

Table 12: Grade 3 reading achievement in Alabama's high implementing school for 2006–2007

Grade 3 measures of achievement			
Year: 2006-2007	DIBELS*	SAT-10*	ARMT*
Profiled RF school	75%	71%	96%
All ARFI schools	66%	52%	77%
All AL schools	67%	68%	85%

* Percentage of students meeting benchmark on DIBELS and meeting third grade proficiency or above on SAT-10 and Alabama Reading and Math Test (ARMT)

Findings from Alabama and California on the relationship *between increased student reading achievement and high implementation* suggest that:

- Where implementation is strong, increases in achievement follow.
- High implementation has a positive and significant impact on English learners' reading achievement.
- High-implementing schools can outperform low-implementing schools on measures of student reading achievement.



The relationship between length of time in high-implementing Reading First schools and high student reading performance

The state evaluation reports provide data to determine the impact of length of time in Reading First on implementation and student achievement. Two states, California and Tennessee, presented their school outcome data by years in the program and by its designation of high- and low-implementation schools.

California

California provides a wealth of data on both implementation and student achievement. Results indicate a steady gain in student achievement for schools participating in Reading First for six years.

Reading First schools have grown significantly since the inception of the program. They have grown significantly more than a statistical control group and significantly more than non-Reading First schools. High-implementing schools have grown significantly more than low-implementing schools. The overall impact of Reading First has been determined through a meta-analysis of all achievement results since 2003. Controlling for school demographic differences, the effect size is 0.082 with a standard error of 0.004. (This is an average standardized beta coefficient calculated using multiple regression analyses, not to be confused with Cohen's d or other measures of effect size.) This 0.082 effect is more than 16 standard errors greater than zero; even two standard errors greater than zero would be sufficient to claim a statistically significant effect with 95 % confidence (Haager et al., 2008, p. 1).

Using the *Reading First Achievement Index* (RFAI), a composite of K–3 measures, student reading achievement has increased “an average of 3.1 points per year, equivalent to 18.6 total points” (Haager et al., 2008, p. 17). The fact that achievement is increasing each year suggests that the length of time in Reading First has a positive impact on student achievement.

Additionally, achievement results from both 4th and 5th grades show that K–3 participation in Reading First benefits students as they move into the upper grades. “A meta-analysis of all grade-related effect sizes shows that the Reading First effect for grades 4 and 5 is on par with the effect for grades 2 and 3” (Haager et al., 2008, p. 17). Thus one can assert that students who participate in Reading First from the early primary years to the intermediate grades continue to benefit as measured by reading achievement, even when they are no longer students in the targeted grades.

The effect size in predicting all possible achievement outcomes provides additional information that supports the positive impact of years in program on student achievement. The Year 6 Evaluation Report addresses this:

The average Reading First (standardized beta) effect size in predicting all possible achievement outcomes since 2003, controlling for starting point and demographic factors, is 0.082 with a standard error of 0.004. This is approximately 16 standard errors higher than zero, where 2 standard errors above zero would be sufficient to claim a statistically significant effect with 95% confidence. This is a conservative estimate of the effect. If the definition of Reading First were to include Years in Program, the Reading First effect would nearly double in size (Haager et al., 2008, p. 18).

The data do, however, also indicate a “plateau effect” when high-implementing schools reach their sixth year of program implementation. This flattening of achievement scores is often observed in “mature programs, leading to the conclusion that length in time of program is not by itself sufficient to support ongoing success (Haager et al., 2008, p. 18).

The “plateau” phenomenon has been reported by other researchers. The greatest increase in level of implementation is expected to occur during the early implementation years, while the rate of implementation slows during later years. This is because as higher levels of implementation are reached, the distance between “high” and “higher” implementation levels are relatively small. The resulting “ceiling effect” has implications for measuring implementation levels (Zhang et al, 2005).

The “plateau” effect is finding its way into the literature on change and instructional reform as well. In the case of Reading First, the expectation of linear gains over time seems appropriate if the goal to have all students reading proficiently by the end of third grade is to be realized. However, recent school improvement studies have recognized that “flat” or “plateau” periods in student performance gains, appearing after initial gains have been made, are normal and should be welcomed as an important and constructive time for reflection and follow-up action.

When performance plateaus appear flat despite considerable effort to improve, one must look deeper in two respects: (1) to see if all the specific ingredients for improvement are actually being worked on, and (2) to realize that the next breakthrough may take additional time for new capacities to kick in. ...Some of these characteristics include having a strong model, the need for greater precision in implementation, and an appreciation of the powerful platform that has been established for going to the next step of improvement. (Fullan, Hill, & Crévola, 2006, p. 7).

These periods of flat performance are actually very important parts of the improvement process—they are the periods which individual teachers consolidate and deepen the knowledge and practices they acquired in earlier stages (of the innovation), in which schools diagnose and identify barriers to the next set of problems and look for the capacity to work on them. (Elmore, 2004, p. 248)

The “plateau” periods are not a warning to shift gears and look for another innovation. They are, instead, a sign that progress has been made and a signal that it is time for reflection, refinement, and re-energizing the innovation to meet evolving needs.

Tennessee

Based on student reading achievement preliminary data and survey results, the number of years a school participated in Reading First had a positive impact on student reading achievement. The Tennessee Reading First 2007 Aggregate Report addresses this connection:

Although individual student data are not yet available, preliminary TCAP results provided evidence as to gains in reading skills for students in the RF schools. Only two of the 74 schools did not make AYP in reading. In addition 63 schools increased the percentage of third graders at proficiency/advanced in reading in 2006–2007 when compared to 2005–2006. Longitudinally, kindergarten, first-, and third-grade students made steady progress on DIBELS.

Teacher perceptions regarding the extent to which student achievement had increased as a consequence of the RF program improved significantly between Years 1 and 4. While 40.1% of Year 1 teachers strongly agreed or agreed that student achievement had been positively impacted by the RF program, 86.3% of teachers in Year 4 strongly agreed or agreed with this statement. After three years of implementation of RF, 84.6% of Round II teachers also strongly agreed or agreed that student achievement had been improved. In addition, at the end of the 2007 year Literacy Leaders in both Round I and II schools indicated that they perceived their schools were progressing toward more fully accomplishing Goal 2 of RF, to ensure that all children can read at grade level or above by the end of third grade. Over 90% of principals also agreed that RF had positively impacted student achievement.

(Grehan, Heegel, Boyraz, & Huang, 2007, p. 3)

Findings from California and Tennessee on the relationship between length of time in high-implementing Reading First schools and high student reading performance suggest that:

- Quantitative data, anecdotal information, and survey results support the notion that length of time in Reading First has a positive impact on increasing student reading achievement, and this impact is greater in high implementing schools.
- After six years of high implementation, schools tend to reach a “plateau” where student achievement scores begin to “flatten.” This occurrence, however, is not an indication of a need for program change, but rather an indication of a need for program refinement.



Concluding remarks

This study attempts to examine the concept of Reading First implementation in selected states. It is based on the premise that implementation is not easily defined nor is it as simple as doing what the grant mandated. It is, instead, a complicated, evolving effort to adhere to a set of practices while working over time to improve those practices to increase student achievement. There has been much success in the Reading First efforts. The states represented in this study have offered perceptions and documentation of their success with the goal of moving all students to proficiency in reading by the end of third grade. While there is still no common agreement in how to define implementation, nor a surefire way to measure it and its effect on student reading achievement, some positive generalizations have emerged from this study and warrant careful consideration.

First, implementation is a complex process that is influenced by a multitude of interrelated factors. Second, the work of implementation is about changing the behavior of teachers and administrators in light of the goal, altering school structures, cultures, and climates in support of the vision and mission, and building the organizational capacity to bring the mission and vision to life. Third, implementing a program like Reading First depends upon well-defined methods of teaching others, ongoing feedback and support from instructional leaders, and finally, a like-minded group focused on implementing the program with fidelity. After fidelity of implementation is assured, the focus turns to making certain that schools are achieving desired results.

Some key conclusions are that:

- Many elements that influence successful implementation go far beyond program compliance factors.
- High-implementing Reading First schools have demonstrated higher student reading achievement gains.
- Length of time in Reading First benefits teachers and students.

At this point, no one is ready to declare victory in terms of fully meeting student achievement goals. We can, however, have confidence that faithful implementation coupled with sufficient time will yield even greater gains in student achievement.

Interview Protocol

INTERVIEW PROTOCOL—State Directors:

State: _____

Person(s) Interviewed: _____

Interviewer: _____

1. How does your State Reading First external evaluation define implementation?
2. How does your State Reading First external evaluation define “high implementation?”
3. How does your State Reading First external evaluation measure implementation?
4. What is the relationship between student success and high implementation?
 - (a) How does your State Reading First external evaluation define student success? (reading proficiency)
 - (b) What percentage of schools are successful high implementers? (if known)
5. How does your State Reading First external evaluation define length of time in the Reading First program?
 - (a) From the beginning of the funding month to final funding month?

Or
 - (b) From the first month of program implementation?
6. What is the relationship between successful high implementation and length of time in Reading First?

References

- Alabama Best Practices Center. (2007). Alabama tackles adolescent literacy. *Working Toward Excellence*, 7(1), 1–5.
- Alabama Best Practices Center. (2004). Can Alabama lead the U.S. in reading achievement? *Working Toward Excellence*, 4(2-3), 1–19.
- Alabama Best Practices Center. (2005). Doing whatever it takes. *Working Toward Excellence*, 5(1), 1–9.
- Alabama Best Practices Center. (2006). Three schools that succeed by doing whatever it takes. *Working Toward Excellence*, 6(1), 1–2.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2, 40.
- Center on Education Policy. (2006). *Keeping watch on Reading First*. Washington, DC: Author.
- Center on Education Policy. (2007). *Reading First: Locally appreciated, nationally troubled*. Washington, DC: Author.
- Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, 72(3), 433–479.
- Durlak, J. & DuPre, E. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327–350.
- Elias, J. M., Zins, E. J., Graczyk, A. P., & Weissberg, P. R. (2003). Implementation, sustainability, and scaling up of social-emotional and academic innovations in public schools. *School Psychology Review*, 32(3), 303–319.
- Elmore, R. F. (2004). *School reform from the inside out: Policy, practice, and performance*. Cambridge, MA: Harvard University Press.
- Emshoff, J. G., Blakely, C., Gottschalk, R., Mayer, J., Davidson, W. S., & Erickson, S. (1987). Innovation in education and criminal justice: Measuring fidelity of implementation and program effectiveness. *Educational Evaluation and Policy Analysis*, 9, 300–311.
- Fixsen, D., Naoom, S. Blasé, K., Friedman, R. & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa: University of South Florida, National Implementation Research Network.
- Fixsen, D., & Paine, S. (2009). *Implementation: Promising practices to sustained results* (Sustainability Series, Number 5). Arlington, VA: RMC Research Corporation.
- Fullan, M., Hill, P., & Crévoila, C.. (2006). *Breakthrough*. Thousand Oaks, CA: Corwin Press.
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., Kyriakidou, O., & Peacock, R. (2005). *Diffusion of innovations in health service organizations: A systematic literature review*. Oxford, England: Blackwell Publishing.
- Grenhan, W. W., Heegel, M., Boyraz, G. & Huang, Y. (2007). *Tennessee Reading First grant formative evaluation 2007 aggregate report*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.
- Guldbansson, K. (2008). *From news to everyday use: The difficult art of implementation*. Östersund, Sweden: Swedish National Institute of Public Health.

- Hall, G. & Hord, S. (2006). *Implementing change: Patterns, principles, and potholes*. Boston: Allyn & Bacon.
- Haager, D., Heimbichner, C., Dhar, R., Moulton, M., & McMillan, S. (2008). *The California Reading First year 6 evaluation report*. Morgan Hill, CA: Educational Data Systems. Retrieved March 18, 2009 from http://www.eddata.com/resources/publications/RF_Evaluation_2007-2008.pdf
- Herlihy, C., Kemple, J., Bloom, H., Zhu, P. & Berlin, G. (2009). *Understanding Reading First: What we know, what we don't, and what's next* (Policy Brief). Washington, DC: MDRC.
- Kurki, A., Aladjem, D. & Carter, K. (2005). *Implementation: Measuring and explaining the fidelity of CSR implementation*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.
- Loucks, S. F. (1983). *Defining fidelity: A cross-study analysis*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.
- MGT of America, Inc. (2007). *Illinois Reading First program evaluation report final 2006–2007*. Tallahassee, FL: Author.
- Miller, S., Swanlund, A., Butler, J., Cisneros, A., Davidson-Gibbs, D. & Lessem, S. (2007). *Wisconsin Reading First: Year 4 2006–2007*. Chicago, IL: Learning Point Associates.
- Moscovitch, E. (2008). *Alabama Reading First outside evaluation 2006–2007 school year*. Gloucester, MA: Cape Ann Economics.
- National Institutes of Health. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: Author. Retrieved May 21, 2009 from <http://www.nichd.nih.gov/publications/nrp/smallbook.cfm>
- O'Donnell, C. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, 78(1), 33–84.
- Stith, S., Pruitt, I., Dees, J., Fronce, M., Green, N., Som, A. et al. (2006). Implementing community-based prevention programming: A review of the literature. *Journal of Primary Prevention*, 27, 599–617.
- United States Department of Education. (2002). *Guidance for the Reading First Program*. Washington, DC: Office of Elementary and Secondary Education, U.S. Department of Education.
- United States Department of Education. (2006). *Reading First Implementation Evaluation: Interim Report*. Washington, DC: Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.
- United States Department of Education. (2008). *Reading First Implementation Evaluation Final Report*. Washington, DC: Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.
- Zhang, Y., Shkolnik, J. & Fashola, O. (2005). *Evaluating the implementation of comprehensive school reform and its impact on growth in student achievement*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.



NATIONAL READING
TECHNICAL ASSISTANCE CENTER

